ARMTRAK: TRACKING ARM POSTURES WITH A SMARTWATCH

BY

SHENG SHEN

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Adviser:

Associate Professor Romit Roy Choudhury

ABSTRACT

We aim to track the 3D posture of the entire arm – both wrist and elbow – using the motion and magnetic sensors on smartwatches. We do not intend to employ machine learning to train the system on a specific set of gestures. Instead, we aim to trace the geometric motion of the arm, which can then be used as a generic platform for gesture-based applications. The problem is challenging because the arm posture is a function of both elbow and shoulder motions, whereas the watch is only a single point of (noisy) measurement from the wrist. Moreover, while other tracking systems (like indoor/outdoor localization) often benefit from maps or landmarks to occasionally reset their estimates, such opportunities are almost absent here.

While this appears to be an under-constrained problem, we find that the pointing direction of the forearm is strongly coupled to the arm's posture. If the gyroscope and compass on the watch can be made to estimate this direction, the 3D search space can become smaller; the IMU sensors can then be applied to mitigate the remaining uncertainty. We leverage this observation to design *ArmTrak*, a system that fuses the IMU sensors and the anatomy of arm joints into a modified hidden Markov model (HMM) to continuously estimate state variables. Using Kinect 2.0 as ground truth, we achieve around 9.2 cm of median error for *free-form* postures; the errors increase to 13.3 cm for a real-time version. We believe this is a step forward in posture tracking, and with some additional work, could become a generic underlay to various practical applications.

To my parents, for their love and support.

ACKNOWLEDGMENTS

I wish to express sincere appreciation to my advisor, Prof. Romit Roy Choudhury, for his assistance in my pursuing the M.S. degree. He has given me invaluable guidance and help with my research work. I'd also like to thank my collaborators, Prof. He Wang and Dr. Mahanth Gowda. Without the fruitful discussions with them and their talent and hardwork, I could not have completed my research. And finally, thanks to my family who have always supported me.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Analytics on human leg motion has fueled an industry on mobile health and well-being. Nowadays, walking, running, biking and various other activities can be recognized from motion sensors embedded in smartphones and wearable devices. Understanding upper limb motion seems like the logical next step, and various research groups/start-ups have already made progress. Authors in [1, 2, 3, 4], for example, have employed various machine learning algorithms to detect meaningful arm and hand gestures – smoking, eating, typing, writing – on wearable wrist bands. Rithmio [5], perhaps the most advanced start-up in this space, is eliminating the need for training, so long as the user performs repetitive tasks, such as bouncing a basketball or exercises in the gym. Finally, for applications requiring full arm posture reconstruction (e.g., golf swing analysis, animation movie characters), today's solutions paste multiple sensors on the arm, or adopt computer vision based analytics [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]. We aim to construct the 3D arm posture using smartwatch sensors alone, and track the posture continuously over time, without any training. In addition, we desire the tracking techniques to be lightweight so they can be amenable to real time applications.

Before further discussion, we briefly clarify the notion of "postures" from "gestures". By posture, we mean the 3D geometric model of the arm. For a fixed shoulder location, arm posture is uniquely defined by three parameters – *elbow location*, *wrist location*, and *wrist rotation*. The wrist rotation captures the rotation of the wrist around the axis of the forearm.[1] A gesture, on the other hand, is a specific sequence of arm postures that carries semantic meaning (somewhat analogous to how words are meaningful sequences of alphabets). Hand gestures typically refer to gestures of the wrist, not

[1]For a fixed elbow and wrist location, the wrist rotation changes the palm's facing direction.

necessarily the arm. We aim to design a system[2] that tracks the entire arm posture in 3D space over time (similar to a Kinect), and is expected to serve as building blocks to any application-defined gesture.

Designing *ArmTrak* entails three key research questions:

(1) The state space of the entire arm is large, meaning that the elbow and wrist could take up many configurations around the body. Without any pre-defined patterns to search for, the arm posture tracking problem translates to a Bayesian tracking problem in continuous space. While tracking is a mature area in signal processing, most of the problems are either guided by good motion models or able to obtain measurements directly from the object of interest. In our case, smartwatch sensors do not offer direct measurements from the elbow, are noisy, and lack models of how the arm is expected to move. To the best of our knowledge (based on literature survey in signal processing, robotics, and mobile computing), this still remains an unaddressed problem.

(2) It is possible that continuous space techniques, such as particle filters or an appropriate variant, map to posture tracking. However, such techniques incur high complexity and latency – when considering scalability to many users, or the real-time requirements of certain applications, the approaches prove prohibitively expensive. Low complexity and fast run-time are important factors for a practical end-to-end system.

(3) The final problem pertains to expressing the arm posture in different coordinate systems. For applications where a user is pointing to a TV to turn it on, it is important to understand the direction of pointing in the global reference frame. For other applications, like golf-swing analysis, hand posture needs to be tracked in the torso's coordinate system. The core problem is rooted in detecting the human's facing direction from the watch sensors. *ArmTrak* must resolve this problem to cater to various application needs.

The perspective we bring to the problem pertains to a synthesis of anatomy, sensor fusion, and Bayesian inference. From the anatomical models of shoulder and elbow joints, we observe that for a given 3D orientation of the wrist (which is estimated via sensor fusion using accelerometer, compass and gyroscope), the space of possible elbow locations is quite constrained. Given that the elbow is also constrained on a sphere around the shoulder point, we can

[2]This work has been published in MobiSys 2016 [24].

further reduce the search space for the elbow – called a *point cloud*. Now, using the (rotation polluted) accelerometer data, we estimate the translational motion of the elbow through a hidden Markov model (HMM) framework, but apply the point cloud as a prior. Once the elbow location is known, the wrist location is computed as a simple shift along the (forearm pointing) direction prescribed by the wrist orientation. To cater to applications, we make a series of optimizations, resulting in an option to prioritize either accuracy or latency. On one extreme, Viterbi decoding yields the best results but after offline processing; on the other extreme, we compute an averaging on the point cloud to operate in real-time. Finally, we use a combination of the watch orientation and the compass to opportunistically estimate the user's facing direction, ultimately yielding the arm posture in the desired coordinate system.

We evaluate *ArmTrak* using Samsung Gear Live smartwatches, with the sensor data processed on the watch (in real time) as well as on the cloud (running MATLAB). Recruited volunteers stand in front of a Kinect 2.0 sensor and perform various kinds of gestures, starting from simple wrist movements all the way to random, free-form arm gestures. The skeletal models from the Kinect serve as ground truth, and we report *ArmTrak*'s accuracy as a function of both the wrist location and elbow location errors. We also report the degradation in our accuracy in exchange for the improvement in latency. On average, our $\langle elbow,\ wrist \rangle$ posture tracking results are $\langle 7.9 \text{ cm},\ 9.2 \text{ cm} \rangle$ respectively in the offline setting, and drop to $\langle 12.0 \text{ cm},\ 13.3 \text{ cm} \rangle$ when performed in the "fast" mode. More importantly, the tracking errors remain bounded over time, allowing for continuous gesture recognition.

Besides what is achievable, we must also discuss the shortcomings of the current system. (1) We believe that ferromagnetic materials in indoor environments can present important ramifications on accuracy; our experiments were performed in our lab with stable magnetic ambience. (2) Our techniques falter when the user performs gestures while on the move – the sensor data from the motion pollutes both posture tracking and facing-direction estimation. (3) Finally, gyroscopes are known to consume energy – we have ignored the energy considerations in developing *ArmTrak*. In view of these capabilities and deficiencies, we summarize our contribution as follows:

- *Using sensor data from smartwatches to track the posture of the entire*

arm. Using observations from anatomical models to constrain the search space for the elbow, a key enabler for 3D posture tracking.

- *Using the accelerometer data as an input to a (modified) hidden Markov model, ultimately tracking the motion of the elbow (and the wrist).* Parameterizing the system to achieve different tradeoffs between accuracy and latency, and offering them as a single knob to application developers.

CHAPTER 2

RELATED WORK

Gesture/posture recognition has been studied from various perspectives. The literature is vast, but we sample the most relevant ones, mainly from computer vision and wearable motion sensors.

Computer vision: Camera data has been used to track and analyze human motion across different granularities [25]. At a lower granularity, humans can be automatically detected [26] and tracked with bounding boxes [27], using the video feeds from cameras. Beyond bounding boxes, human activity can also be recognized from camera data via machine learning[28, 29]. At a higher granularity, pose estimation is a classical problem in human motion analysis, where the common approach is applying probabilistic models on the static RGB image or video sequence [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]. More recently, depth information has also been leveraged in the pose estimation solution landscape [18, 19]. For instance, Microsoft Kinect [18] fuses RGB and depth image to track the locations of the human's joints for video gaming. One of the key differentiators between vision and sensing based approaches is that vision must estimate 3D motion from the 2D views – a challenging task. However, vision benefits from knowing the pixel locations far more precisely compared to the noise from sensor hardware.

Wearable motion sensors: Previous research has shown that embedded motion sensors on wearable devices can be used for human activity recognition [30]. Industry on mobile health and well-being uses these sensors to recognize a user's leg motion including walking and running [31, 32]. To measure meal intake, Bite Counter [2] uses a watch-like device with a gyroscope to detect and record when an individual has taken a bite of food. RisQ [1] leverages motion sensors on wristband to recognize smoking gestures. MoLe [3] analyzes motion data of smartwatches from typing activity to infer what the user has typed. Xu et al. [4] classified hand/finger gestures and written characters from smartwatch motion sensor data. However, all these motion

analysis systems are designed for recognizing specific pre-defined motion patterns, as opposed to blind estimation of free-form postures. Similarly, authors in [33, 34] tried to reconstruct full body motion from multiple wearable devices by comparing accelerometer data with those generated from motion capture databases. However, the reconstruction relies heavily on the similarity of training and testing accelerometer data and the disparity between different motion classes inside training databases, and as a result, neither can they track free-form arm motion.

Zhou et al. [20, 21], Cutti et al. [22], and El-Gohary et al. [23] studied general upper limb movement tracking using motion sensors. However, they require users to be instrumented with multiple sensors on the arm. Perhaps the closest to our work is [35], where authors claimed to be able to track the upper limb by only mounting motion sensors on the wrist. However, the system is only evaluated on one subject, moving his arm up-to-down in a plane perpendicular to the ground. The same gesture is repeated constantly and lasted less than 15 s. As a follow-up to this work, the authors published a subsequent paper with multiple sensors on the arm to scale to a larger vocabulary of gestures [20, 21]. In contrast, our system is tested with free-form motion, has been tested up to 3 mins without signs of divergence, and has demonstrated robustness to all eight test users. We believe this is an improvement over the state of the art.

IR technology (Vicon [36], Optitrack [37]) is popularly used for gesture tracking and gait analysis, where the entire skeletal motion can be reconstructed with mm level accuracy based on reflections of IR signals from IR markers pasted on the body. Besides being expensive, they require instrumentation of humans and environment with multiple markers and IR cameras. **Wireless sensing** such as Witrack [38], WiSee [39], and RF-Capture [40] tracks motion of body parts by analyzing body radio reflections. However, tracking is only effective when the body-parts are moving in the direction of the antenna array. Ubiquitous tracking of wrist and elbow locations is hard. While still not perfect, wireless sensing works behind walls and static occlusions. However the environment still needs to be instrumented and the tracking range is limited. **Light**-based systems [41] reconstruct user skeleton from shadows, but they also require instrumentation on the environment.

CHAPTER 3

PROBLEM SETUP

3.1 Torso Coordinate System

We will define the posture of the arm, and its motion, in the torso coordinate system (Figure 3.1). In this system, the left shoulder will serve as the origin, and the plane of the user's torso (i.e., the chest) will serve as the YZ plane. The X-axis will be the line emanating from the left shoulder in the frontward direction, perpendicular to the torso. The Kinect also models its skeleton tracking data in a similar coordinate system – since we use the Kinect as ground truth, aligning our coordinate system with Kinect simplifies our evaluation process.

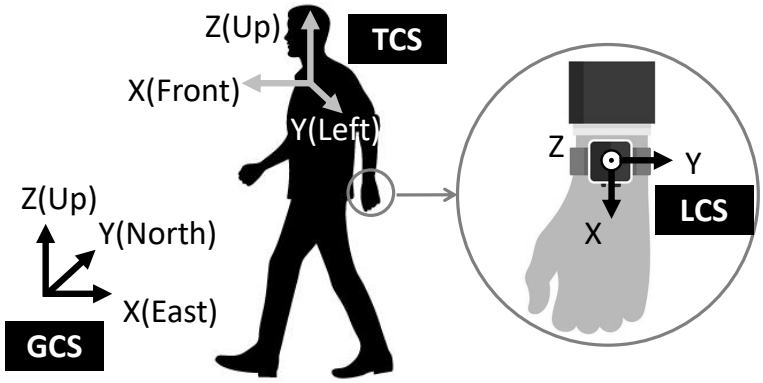

Figure 3.1: Torso coordinate system (TCS), global coordinate system (GCS) and local coordinate system (LCS).

The torso coordinate system (TCS) is desirable to most applications, although some need the arm posture to be expressed in a global (North-East) coordinate system (GCS). For instance, analysis of gym exercises, golf swing analysis, smoking recognition, etc., can all be performed in the TCS framework. However, when controlling devices in a room (e.g., pointing to a TV to turn it on), the posture of the arm needs to be modeled in global coordi-

nates. The compass on the watch offers the necessary information to estimate postures in GCS, however, its translation to TCS requires knowledge of the user's facing direction. We will develop the overall *ArmTrak* system assuming knowledge of the facing direction, and then relax the assumption through a facing-direction estimator.

3.2 Wrist/Elbow Orientation and Location

Figure 3.2 shows the X-, Y-, and Z-axes of a smartwatch when a user wears it on his or her left wrist. These axes, that are local to the watch's coordinate system, can easily be expressed as vectors in the torso coordinate system, denoted as $\vec{X}_t$, $\vec{Y}_t$, and $\vec{Z}_t$ (the subscript means the vector changes over time as arm moves). We define the "orientation" of the watch (same as the orientation of the wrist) as this tuple: $\langle \vec{X}_t, \vec{Y}_t, \vec{Z}_t \rangle$. $\vec{X}_t$ always aligns with the pointing direction of the forearm in 3D space, and for a fixed forearm pointing direction, both $\vec{Y}_t$ and $\vec{Z}_t$ change along with the rotation of the forearm around the X-axis.

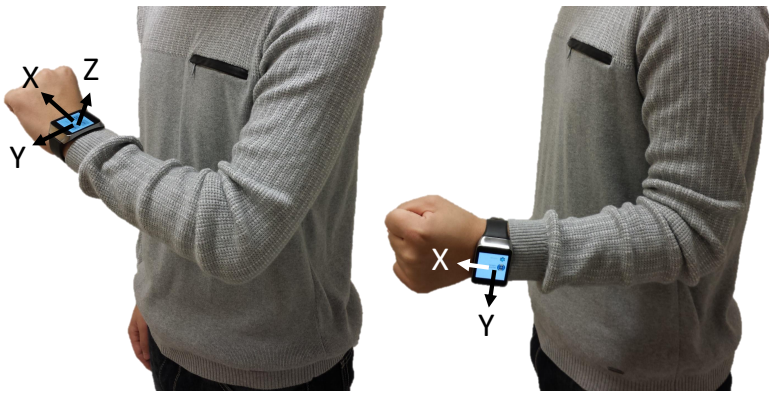

Figure 3.2: Smartwatch orientation changes with the pointing direction of the hand – the orientation is measured in the torso's reference frame.

Similarly, location of the elbow or the wrist (same as the location of the watch) can also be expressed as a separate 3D tuple, $\langle x_t, y_t, z_t \rangle$, in the torso coordinate system (TCS). Once the wrist orientation is known, the elbow location and the wrist location are simply a *static shift* of each other, along the positive or negative forearm pointing direction (wrist's $\vec{X}_t$ orientation). The static shift is the length of the forearm, and thus needs to be known only once. With this, the posture of the entire arm can be determined in

the torso coordinate system. The natural question is: How can the arm posture be tracked over time? Actually, we can ask a simpler sub-question: Can the wrist location even be tracked over time? We make a few relevant observations next.

3.3 Noise: The Fundamental Problem

From basic physics, any motion of a body can be decomposed into translational and rotational motion. Thus, when a wrist moves from point A to a very close point, B, one can model this as a translational motion of the watch from A to B, followed by a change in orientation to reflect the orientation at B. Now, assume that the initial location and orientation are known at point A, and the accelerometer and gyroscope are super accurate. Then, the gyroscope can measure the angular velocity around each of the $\vec{X_t}$, $\vec{Y_t}$, and $\vec{Z_t}$ directions, and precisely estimate the orientation at point B. The accelerometer, on the other hand, measures a combination of translational and rotational motion (hence, plain double integration will not work). Instead, the double integration can be performed at infinitesimally small time steps, and after each step, the orientation of the device can be updated based on the corresponding gyroscope data. In other words, the system will be able to compute the linear displacements and rotations at extremely fine granularity, and concatenating them should result in perfect tracking.

Of course, the above is true under the assumption of perfect IMU sensors. With noisy sensors, we implemented the same algorithm (and appropriately subtracted gravity) to quantify the extent of divergence. Figure 3.3 shows the results – the orientation divergence is somewhat reasonable,[1] but the translation error is excessive, more than 100 meters within 1 minute. In other words, deterministic techniques will always be affected by the randomness of noise; stochastic inference techniques are likely to be the appropriate approach.

[1]We cannot measure all three dimensions of orientation using Kinect, hence plot the error from two dimensions, calculated as the angular difference between the forearm's pointing direction and the ground truth.

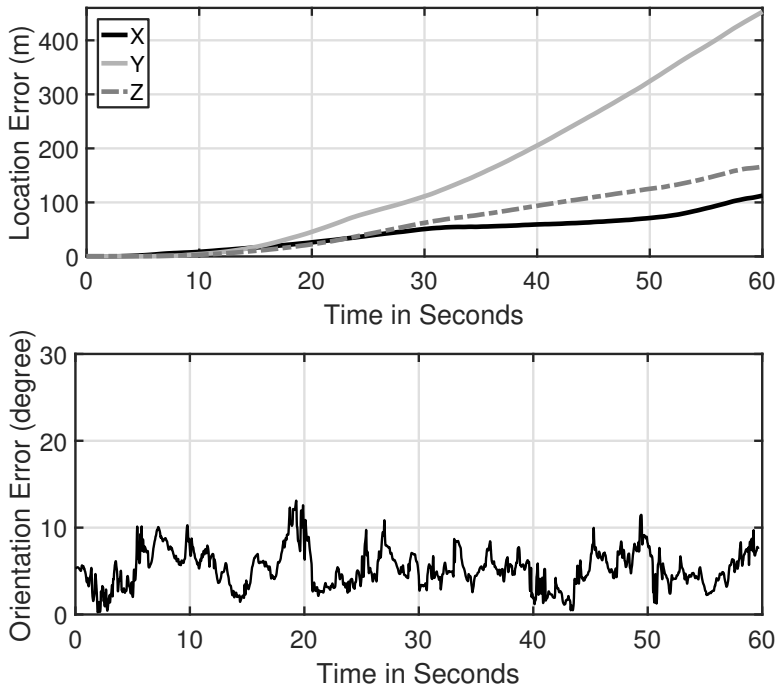

Figure 3.3: (a) Wrist location error diverges with double integral. (b) Wrist orientation error remains small over time.

3.4 An Estimation Problem: Particle Filter

The problem we are facing is obviously not new – robotics and signal processing researchers routinely face and solve these kinds of estimation problems (also called filtering). Briefly, the *state* of the object is modeled as variables and the range of these variables together describe the *space* in which the object can exist. Then, based on some model of how the object is expected to move, and what the data reveals about its actual motion, these estimation algorithms compute the most likely state of the object. We implemented a particle filter, one of the popular state estimation algorithms in continuous space. However, given that the accelerometer and gyroscope data are differentials of location and orientation, the state of particles had to be defined with many variables to capture the entire arm posture. This resulted in a high-dimensional system and the estimator could hardly converge. We aborted the effort and focused on reducing the state space of the system for good tracking accuracy.

CHAPTER 4

OPPORTUNITY AND VALIDATION

Following the failure of the particle filter, we focused on opportunities to reduce the state space of the system, i.e., constraining the possible postures of the arm. This seemed intuitive, i.e., since the arm joints have limits in their *range of motion* (RoM) [42, 43], they should constrain the arm postures as well. In exploring the arm joints and measurement data, we made the following empirical observation. Assume the shoulder location is fixed. *It appeared that for a fixed wrist orientation, the possible space of wrist locations is quite limited.* In other words, if one moves his or her wrist around without changing the wrist orientation, there are not many locations to which that person can take the wrist.

To understand this intuition, let us first assume that we keep the elbow location fixed. Consider how the motion of the forearm will influence the wrist location and orientation. Since the elbow does not move, any forearm motion will change the wrist's orientation, no matter it is the twist of the wrist/forearm (which will change $\vec{Y_t}$ and $\vec{Z_t}$ of orientation) or the rotational motion around the elbow (which will change $\vec{X_t}$ of orientation and also change the wrist's location). Conversely, for a given wrist orientation, only one wrist location is possible (as it has to be along the $\vec{X_t}$ direction emanating from the elbow), under these artificial assumptions.

Of course, once the elbow starts moving, the wrist can move to multiple locations while preserving the same orientation. However, the elbow can only move on a sphere around the shoulder, and the forearm's ability to twist is relatively limited. This suggests that for a given wrist orientation, the possible space of wrist locations may be reasonably restricted. The space will also vary across orientations, i.e., some wrist orientations will allow the wrist to move to more locations than others.

4.1 Preliminary Validation

As preliminary validation, we visualized the space of wrist locations for some wrist orientations. Figure 4.1 shows five example wrist orientations, along with the corresponding wrist and elbow location point clouds, marked in (light) green and (dark) red, respectively. The findings exhibit promise – the wrist location space is indeed a small fraction of the entire 3D space around the shoulder. Also, the elbow and wrist point clouds exhibit a 1:1 mapping, since they are simply a static shift of one another.

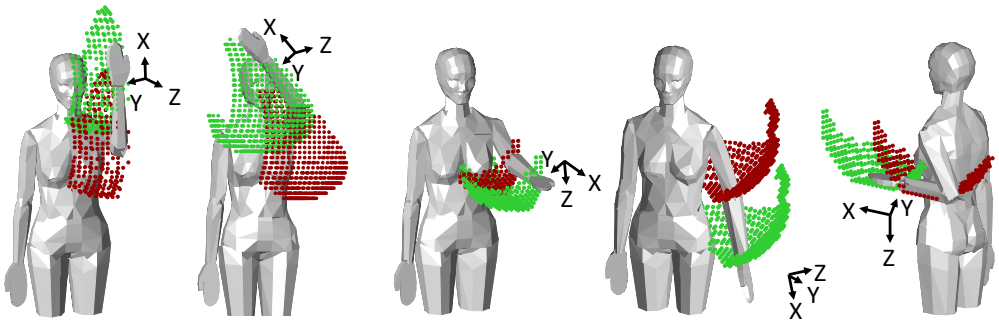

Figure 4.1: Five different watch orientations and the corresponding wrist and elbow point clouds (shown in light green and dark red, respectively). The point clouds (essentially the space of feasible wrist and elbow locations for the given orientation) are relatively small and narrow down the uncertainty of the user's arm posture.

4.2 Formalizing through Arm Models

Figure 4.1 indicates the opportunity, but we need to generalize our observation. Therefore, we derived models from human arm kinematics, and reorganized them to formally express the relationship of wrist orientation, wrist location and elbow location. We describe the models here, followed by a quantification of state space reduction.

In robotics, a human arm is often modeled using 7 rotational degrees of freedom (DoF) [44] – 3 for the shoulder, 2 for the elbow, and 2 for the wrist. Since the watch is worn on the forearm near the wrist, the DoFs of the wrist are not manifested in the watch's sensor data. The remaining 5 DoFs define the state of the watch, as shown in Figure 4.2. When these five values are

combined with the *known* lengths of the upper arm and forearm, the watch's location and orientation can be estimated uniquely.

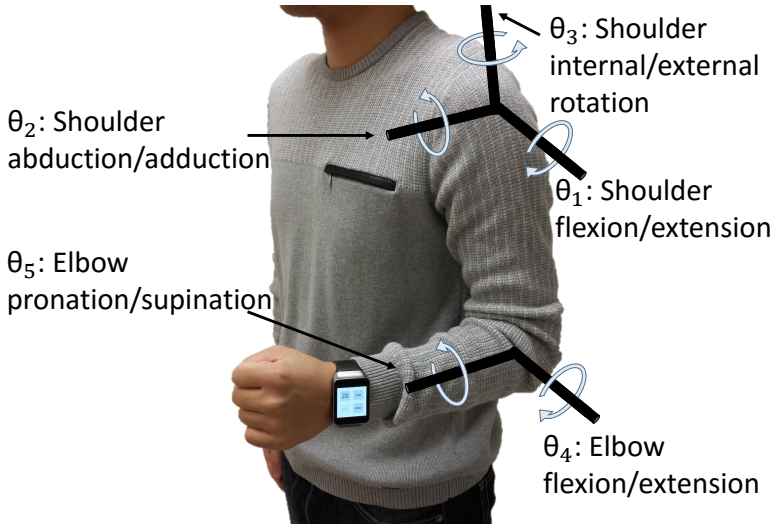

Figure 4.2: 5-DoF arm model showing the possible angular rotations.

Modeling this mathematically, let θ_1, θ_2, θ_3, θ_4 and θ_5 denote the 5 DoFs; let l_u and l_f denote the lengths of the upper arm and forearm. Using these, the Denavit-Hartenberg transformation [45] actually outputs the posture of the entire arm. For example, the elbow location is a function of θ_1 and θ_2, and can be expressed as:

$$\text{loc}_{\text{elbow}} = f(\theta_1, \theta_2) = l_u \begin{pmatrix} \cos(\theta_2)\ \sin(\theta_1) \\ \sin(\theta_2) \\ -\cos(\theta_1)\ \cos(\theta_2) \end{pmatrix} \tag{4.1}$$

and it satisfies

$$\| \text{loc}_{\text{elbow}} \| = l_u \tag{4.2}$$

Similarly, the wrist's relative location to the elbow can also be computed as

$$\text{loc}_{\text{wrist-to-elbow}} = g(\theta_1, \theta_2, \theta_3, \theta_4) \tag{4.3}$$

where the function $g()$ is a long equation omitted in the interest of space, but it of course satisfies

$$\| \text{loc}_{\text{wrist-to-elbow}} \| = l_f \tag{4.4}$$

Thus, the wrist's absolute location can be written as the vectorial addition of the elbow location and the wrist's relative location (to the elbow).

$$\text{loc}_{\text{wrist}} = \text{loc}_{\text{elbow}} + \text{loc}_{\text{wrist}-\text{to}-\text{elbow}} \tag{4.5}$$

Like location, the orientation of the wrist, expressed in the form of rotation matrix, can also be computed through a rotational function on the 5 θs (the function $h()$ omitted in the interest of space).

$$\text{Rot}_{\text{watch}} = h(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5) \tag{4.6}$$

In summary, knowing these five values for θs can solve the entire state of the arm posture, i.e., wrist orientation, wrist location and elbow location.

4.2.1 Mapping Orientation to Point Cloud

For a given watch orientation, we intend to map it to the elbow and wrist's location point clouds. We derive the mapping to the elbow first, because it lies on a sphere around the shoulder which will later make the model mathematically easier. The translation from the elbow to the wrist will be a static shift.

Now, to derive the elbow's point cloud, we first referred to some medical papers [42, 43, 46] and summarized the average range of motion (ROM) for each joint angle in Table 4.1. Here $\theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5 = 0^\circ$ refers to the posture where the left arm is in free-fall on the left side of the torso, with the palm facing front.

Table 4.1: Range of motions for each joint angle.

Joint Angle	Min. Value	Max. Value
θ_1	-60°	180°
θ_2	-40°	120°
θ_3	-30°	120°
θ_4	0°	150°
θ_5	0°	180°

Then, for each watch orientation, we find all combinations of $\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ within the ROM that can generate that orientation according to Equation

(4.6). Each combination will map to one elbow location according to Equation (4.1). Thus, we obtain a mapping from $\text{Rot}_{\text{watch}}$ to possible values of $\text{loc}_{\text{elbow}}$. We can also derive the mapping from $\text{Rot}_{\text{watch}}$ to possible values of $\text{loc}_{\text{wrist}}$ easily, because for each $\text{Rot}_{\text{watch}}$, possible wrist locations are simply a shift of possible elbow locations, along the forearm's pointing direction – shown in Equation (4.7).

$$\text{loc}_{\text{wrist-to-elbow}} = \text{Rot}_{\text{watch}} \begin{pmatrix} l_f \\ 0 \\ 0 \end{pmatrix} \tag{4.7}$$

Algorithm 1 presents the pseudo code.

Algorithm 1 Watch Orientation to Point Cloud Mapping

ElbowPointCloud = Empty Dictionary
WristPointCloud = Empty Dictionary
for all $\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\} \in$ ROM **do**
  $\text{loc}_{\text{elbow}} = f(\theta_1, \theta_2)$
  $\text{Rot}_{\text{watch}} = h(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$
  $\text{loc}_{\text{wrist-to-elbow}} = \text{Rot}_{\text{watch}}(t) \begin{pmatrix} l_f \\ 0 \\ 0 \end{pmatrix}$
  $\text{loc}_{\text{wrist}} = \text{loc}_{\text{elbow}} + \text{loc}_{\text{wrist-to-elbow}}$
  $\text{ElbowPointCloud}[\text{Rot}_{\text{watch}}].\text{Add}(\text{loc}_{\text{elbow}})$
  $\text{WristPointCloud}[\text{Rot}_{\text{watch}}].\text{Add}(\text{loc}_{\text{wrist}})$
end for

To quantify the reduction in uncertainty due to this mapping, Figure 4.3 plots the CDF of the elbow's point cloud, as a fraction of the surface area of the sphere around the shoulder. In 90% of the cases, the elbow can only reach $\frac{1}{4}$ of the whole sphere area, and the median fraction is 8.3%. Of course, the fraction can be further reduced if we utilize the fact that these five DoFs are not entirely independent and thus model their RoMs jointly (instead of setting an upper/lower bound for each of the joint angle). We leave this optimization to future work.

If the mapped point cloud is moderately accurate (assuming that the orientation estimation is reasonably error-free), the next step is to leverage the point cloud in a state estimation framework. This motivates a discrete space hidden Markov model, with the elbow's point cloud as the prior. We describe

these techniques next, as a part of a full posture recognition system.

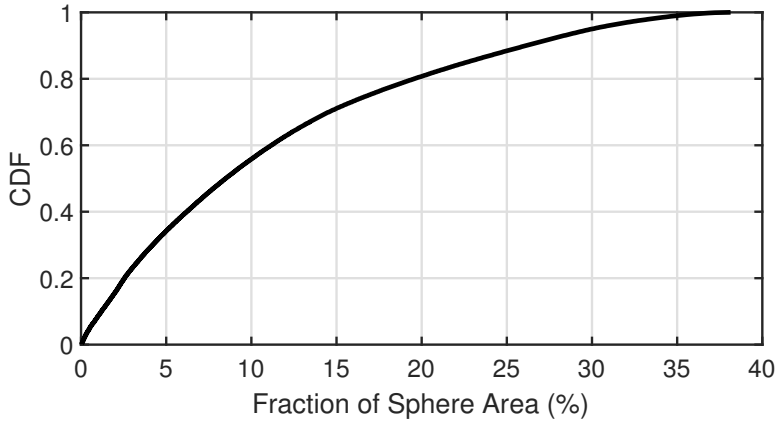

Figure 4.3: Elbow location subspace as a fraction of the sphere area (around the shoulder) – the state space is certainly smaller.

CHAPTER 5

ARCHITECTURE

Figure 5.1 illustrates *ArmTrak*'s overall architecture. The raw sensor data from the smartwatch – composed of the accelerometer, gyroscope, and compass samples – are passed through an Orientation Estimation Module (OEM). This module computes the watch's orientation in the earth's coordinate system (ECS), using a borrowed technique called A^3 from MobiCom 2014 [47]. Since the user's facing direction is unknown, the transformation between ECS to TCS is still unknown. The Facing Direction Module (FDM) scans the sensor data stream and opportunistically recognizes samples that reveal the facing direction. The orientation is now transformed to TCS and forwarded to the Orientation to Point Cloud Module (OPM).

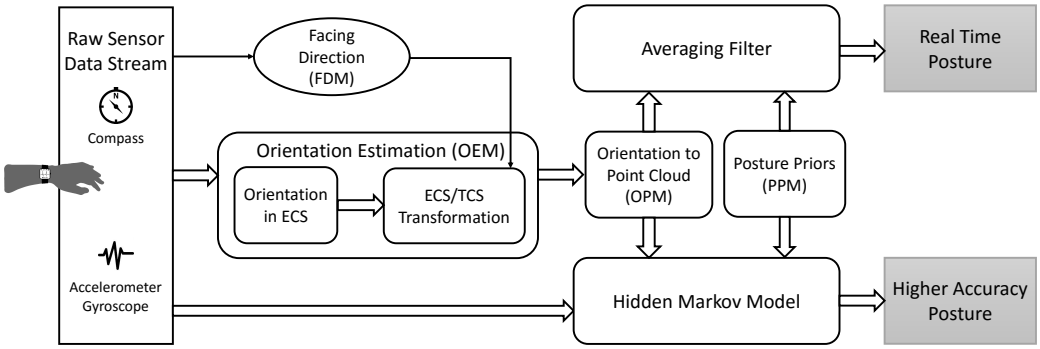

Figure 5.1: System architecture. The raw sensor data is processed to obtain watch orientation and possible arm postures, and together with posture priors, they are sent to two different filters to obtain different posture estimations.

OPM consists of a pre-loaded mapping between orientation and point clouds (the mapping derived from our arm-joint models described in Section 4.2). Using the incoming orientation as an index, OPM outputs the corresponding point cloud. Not all candidates may be equally likely, therefore, a separate Posture Priors Module (PPM) analyzes general human arm

motions and extracts priors. This information is used to bias the posture estimation process towards the sequences that are more likely in humans.

For applications that require high accuracy in arm posture estimation, the outputs of both OPM and PPM are forwarded to a Hidden Markov Model (HMM), along with the raw sensor data. The HMM observes sequences of data and estimates the most likely arm posture sequence. The outputs are favorable to applications that need higher accuracy and can tolerate latency in several seconds (even though the HMM has been carefully optimized for lower complexity). However, if some applications require a real-time arm posture, the outputs of OPM and PPM are sent together into an averaging filter. This filter computes a weighted average of all candidate arm postures for the current orientation, where the weights are guided by the priors. This serves as a faster but less accurate arm posture tracker.

5.1 Design for Higher-Accuracy Postures

We first design for accuracy without latency considerations. As described in Section 4.2, once the orientation of the watch is known, posture estimation boils down to an elbow tracking problem. If the elbow can be tracked, the wrist location can be computed as a static shift, yielding the complete arm posture. The resources we have (in addition to sensor data) are twofold: (1) reasonable estimates of orientation, even though not precise, and (2) point cloud of all possible elbow locations, for a given orientation. The question then is: *At any given time, where is the elbow in the point cloud?*

Recall that the point cloud is often quite small – on average, it covers less than 10% of the sphere around the shoulder (Figure 4.3). Simply using the average location of the point cloud could result in a reasonably good estimate of the elbow location. However, in testing this simple averaging method with various kinds of gestures, we found much room for improvement. For instance, consider the punching gesture in Figure 5.2 – the forearm moves forward and backward while the orientation remains the same. The averaging scheme always shows the center of the point cloud (see Figure 5.3) since the point cloud remains almost the same for the entire gesture.

The room for improvement (over simple averaging) arrives from using the smartwatch sensors as an estimate of the elbow's motion. In this specific

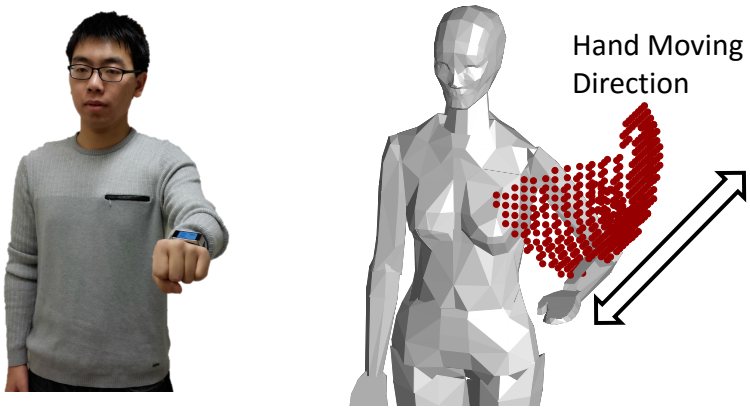

Figure 5.2: Punching: (a) video frame; (b) point cloud remains almost the same during punching.

punching case, the accelerometer data from the watch can be used to estimate the elbow's acceleration (in a straight line), which in turn can be converted to the wrist. In general, however, this is more complicated since the elbow will also experience rotational motion – in such cases, its acceleration has to be computed through a fusion of the gyroscope and accelerometer. To this end, we first introduce techniques to estimate acceleration, and then combine this elbow acceleration with the point cloud constraints to estimate elbow location.

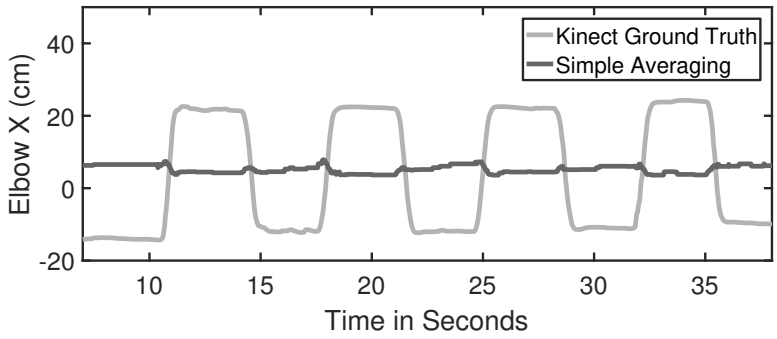

Figure 5.3: X coordinate of elbow location. During punching gesture, the orientation of the watch remains almost the same; therefore the point cloud does not change much over time. As a result, averaging the point cloud cannot follow the location of the elbow.

5.1.1 Estimate Elbow Acceleration

For any physical object, the acceleration is simply the second derivative of the location time series. For the elbow, we have

$$\text{accel}_{\text{e}}(t) = \frac{\text{d}^2}{\text{d}t^2}\text{loc}_{\text{e}}(t) \tag{5.1}$$

Applying Equation (4.5), we have:

$$\begin{aligned}\text{accel}_{\text{e}}(t) &= \frac{\text{d}^2}{\text{d}t^2}\left[\text{loc}_{\text{w}}(t) - \text{loc}_{\text{we}}(t)\right] &(5.2)\\ &= \frac{\text{d}^2}{\text{d}t^2}\text{loc}_{\text{w}}(t) - \frac{\text{d}^2}{\text{d}t^2}\text{loc}_{\text{we}}(t) &(5.3)\end{aligned}$$

The first term in Equation (5.3), $\frac{\text{d}^2}{\text{d}t^2}\text{loc}_{\text{w}}(t)$, is simply wrist's acceleration in torso coordinate system, which we can get by projecting watch's accelerometer readings into torso coordinate system using estimated watch's orientation, $\text{Rot}_{\text{watch}}$:

$$\frac{\text{d}^2}{\text{d}t^2}\text{loc}_{\text{w}}(t) = \text{Rot}_{\text{watch}}(t)\ \text{accel}_{\text{watch}}(t) \tag{5.4}$$

The second term in Equation (5.3), $\frac{\text{d}^2}{\text{d}t^2}\text{loc}_{\text{we}}(t)$, is the acceleration caused by wrist's relative motion to the elbow. According to Equation (4.7), we can express this as:

$$\frac{\text{d}^2}{\text{d}t^2}\text{loc}_{\text{we}}(t) = \frac{\text{d}^2}{\text{d}t^2}\left[\text{Rot}_{\text{watch}}(t)\begin{pmatrix} l_f \\ 0 \\ 0 \end{pmatrix}\right] \tag{5.5}$$

Now, combining Equation (5.4) and Equation (5.5), we can re-write Equation (5.3) as

$$\begin{aligned}\text{accel}_{\text{e}}(t) = &\ \text{Rot}_{\text{watch}}(t)\ \text{accel}_{\text{watch}}(t)\\ &- \frac{\text{d}^2}{\text{d}t^2}\left[\text{Rot}_{\text{watch}}(t)\begin{pmatrix} l_f \\ 0 \\ 0 \end{pmatrix}\right] &(5.6)\end{aligned}$$

Equation (5.6) shows that given watch's accelerometer data and our estimation on the watch's orientation, the elbow acceleration can be inferred.

5.1.2 Estimate Elbow Location

Given the measured elbow acceleration, as well as the point clouds (on the sphere) on which the elbow must be located, we now ask the following question: *Which sequence of elbow locations best matches the measured elbow acceleration?* To intuitively understand this problem, we assume there are N locations that the elbow can possibly reach – this can be viewed as the union of all point clouds for the elbow. Assume the motion sequence contains T time steps. Since the elbow can be at any of the N locations at a given step, the possible number of sequences is N^T, which is the search space for tracking the elbow. One of the sequences is optimal and our goal is to find this sequence. Hidden Markov Models (HMM) are well suited to solve this problem due to its dynamic programming construction for efficiently searching the state space.

5.1.3 Modified HMM for Elbow Tracking

If we group three locations (at three consecutive time steps) as the state of the elbow, then elbow acceleration can be encoded in one state. By using the estimated acceleration as observation and properly designing transition probabilities, finding the best elbow location sequence reduces to the Viterbi algorithms (solvable in polynomial time). Viterbi decoding has a time complexity of $O(|S|^2T)$, where $|S|$ is the state space size and T is the total number of time steps. In the above HMM formulation, the state space size is N^3, since each state is a location triple. As a result, the time complexity is $O(N^6T)$. In trying to reduce the running time, we reorganized the state definitions. Specifically, we apply the continuity constraint, move the emission probability into the transition probability, and $|S|$ can be greatly reduced by allowing each state to contain only two locations. The time complexity is reduced to $O(N^3T)$. The details are presented below.

For the ease of the description, we assign each possible elbow location on the sphere a location ID, ranging from 1 to N. We also denote the T time steps of the motion sequence as $t_1, t_2, ..., t_T$, and the time step length as ΔT.

• **State definition**: Each state is defined as a pair of elbow locations:

$$\text{state}_i = \langle \text{loc}_\text{e}^{(i_1)}, \text{loc}_\text{e}^{(i_2)} \rangle \tag{5.7}$$

where $loc_e^{(i_1)}$ and $loc_e^{(i_2)}$ are the locations with location IDs i_1 and i_2, which together uniquely define the i-th state.

In our HMM formulation, we use a state to represent the elbow's previous location and current location. In this representation, if the state at time t_k is state_i, it means that the elbow location is $loc_e^{(i_1)}$ at t_{k-1} and $loc_e^{(i_2)}$ at t_k.

At first glance, the size of state space is N^2. However, observe that the human's elbow movement is limited to a maximum speed, thus given a small time step, the elbow can only move within a small range. We can actually eliminate those states whose location pairs are separated by larger than this range. In this way, the size of state space is reduced to αN^2, where α is much smaller than 1.

• **Prior probability**: Since we do not know the initial elbow location, we set the prior probability to be uniform:

$$\Pi(\text{state}_i) = \frac{1}{\alpha N^2} \qquad \text{for any } i \tag{5.8}$$

• **Transition probability**: From current time t_k to next time t_{k+1}, the transition probability from state_i to state_j, $Pr(\text{state}_j \mid \text{state}_i; t_k, t_{k+1})$, contains three terms.

First, since the elbow trajectory is continuous, state_i and state_j must share the same location at the common time step t_k. This continuity constraint actually helps reduce the time complexity from $O(N^4T)$ to $O(N^3T)$.

$$\begin{aligned} \text{state}_i &= \langle \text{loc}_\text{e}{}^{(i_1)}, \text{loc}_\text{e}{}^{(i_2)} \rangle \\ \text{state}_j &= \langle \text{loc}_\text{e}{}^{(j_1)}, \text{loc}_\text{e}{}^{(j_2)} \rangle \\ \text{loc}_\text{e}{}^{(i_2)} &= \text{loc}_\text{e}{}^{(j_1)} \end{aligned} \tag{5.9}$$

We can express this limitation as an indicator function:

$$Pr_1 = I_{\ \text{loc}_\text{e}{}^{(i_2)} \ = \ \text{loc}_\text{e}{}^{(j_1)}} \tag{5.10}$$

Second, instead of using measured elbow acceleration as an observation in the location triple, here we directly model that probability into the transition probability between two location tuples. To be more specific, we can calculate the speed encoded in each state and derive acceleration using the speed of

the two states:

$$\begin{aligned} \text{velocity}_j &= \frac{\text{loc}_\text{e}^{(j_2)} - \text{loc}_\text{e}^{(j_1)}}{\Delta T} \\ \text{velocity}_i &= \frac{\text{loc}_\text{e}^{(i_2)} - \text{loc}_\text{e}^{(i_1)}}{\Delta T} \\ \text{accel}_{i,j} &= \frac{\text{velocity}_j - \text{velocity}_i}{\Delta T} \end{aligned} \tag{5.11}$$

This acceleration, $\text{accel}_{i,j}$, is expected to be close to our observed acceleration $\text{accel}_{\text{observe}}(t_k)$. We assume that the error distribution of the observed acceleration is a zero mean Gaussian distribution with a standard deviation of σ_{accel}. Therefore, we have:

$$Pr_2 = \frac{1}{\sqrt{2\pi}\sigma_{\text{accel}}} e^{(\text{accel}_{i,j} - \text{accel}_{\text{observe}}(t_k))^2/(2\sigma_{\text{accel}}^2)} \tag{5.12}$$

Third, in the new state_j, the elbow location $\text{loc}_\text{e}^{(j_2)}$ must be inside the point cloud inferred at time t_{k+1}.

$$Pr_3 = I_{\ \text{loc}_\text{e}^{(j_2)} \in \text{PointCloud}_{t_{k+1}}} \tag{5.13}$$

In sum, the transition probability is the product of these three probabilities:

$$Pr(\text{state}_j \mid \text{state}_i; t_k, t_{k+1}) = Pr_1 Pr_2 Pr_3 \tag{5.14}$$

• **Emission probability**: Since we have already integrated the observed acceleration into transition probability, emission probability is simply set as 1. Thus, we only use the output of Viterbi decoding – it is a sequence of states and the second element of each state is the estimated elbow location. Figure 5.4 shows the improved elbow tracking results (in this toy punching case) using HMM.

• **Facing Direction**: We opportunistically sense the facing direction when the user's hand is in the vertical free-fall posture, or swinging through this position perhaps while walking. At this point, the X-axis of the gyroscope is exactly pointed in gravity's direction, implying that the negative Y is the facing direction. We also observe that the hand swings to a small degree in this position and we utilize this to gain confidence.

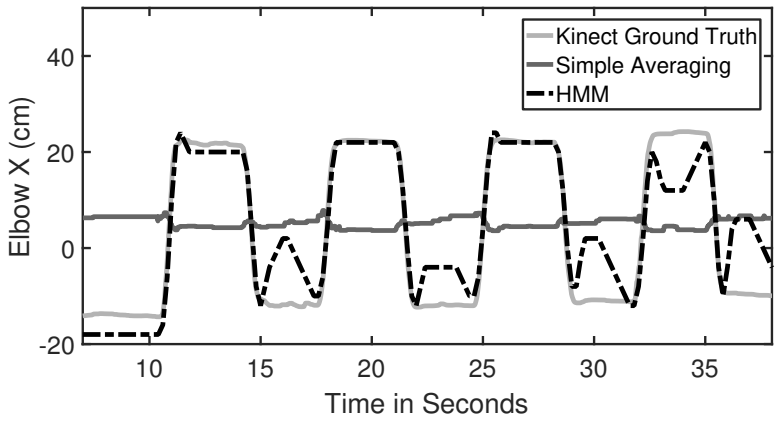

Figure 5.4: X coordinate of elbow location. Results from HMM better capture the location of the elbow.

5.2 Designing for Fast Posture Tracking

For the applications on smartwatches that require instantaneous posture information, the tracking algorithm must be lightweight in order to fit into the watch's limited computing power. As hinted earlier, we adopt a simple weighted averaging filter. This filter accepts (1) the point cloud of arm postures and (2) the corresponding prior from past measurements, and then outputs a weighted average of all the points in the cloud. The weights are determined by the probability density of each location – the more common postures will naturally bias the estimates. The weighted average is expressed as:

$$\mathrm{loc_e} = \sum_i \mathrm{loc_e^i} \frac{Pr(\mathrm{loc_e^i})}{\sum Pr(\mathrm{loc_e^i})} \tag{5.15}$$

$$\mathrm{loc_w} = \mathrm{loc_e} + \mathrm{Rot_{watch}}(t) \begin{pmatrix} l_f \\ 0 \\ 0 \end{pmatrix} \tag{5.16}$$

As an outcome of averaging, the results are naturally quite smooth. Also, the averaged 3D location (for each point cloud) can be stored in a lookup table, indexed by the orientation corresponding to the point cloud. For a given orientation from the Orientation Estimation Module, the smartwatch simply looks up the elbow location from this table, computes the corresponding

wrist location, and outputs the posture. Both memory and CPU footprint is marginal.

CHAPTER 6

IMPLEMENTATION AND EVALUATION

This chapter discusses the experiment methodology and performance results of *ArmTrak*.

6.1 Implementation

ArmTrak is implemented on the Samsung Gear Live smartwatch using JAVA as the programming platform. The accelerometer and gyroscope are both sampled at 200 Hz and the magnetic field sensor is sampled at 100 Hz. The smartwatch runs the lightweight real-time version of *ArmTrak* to report instantaneous arm posture. The sensor data and lightweight arm posture estimates are stored locally and transferred to the *ArmTrak* server for analysis. The server side code is written in MATLAB and implements the full version of *ArmTrak* to provide offline, higher-accuracy, posture estimates.

6.2 Methodology

We recruited eight volunteers, including six males and two females, for our experimentation. The volunteers were asked to wear a Gear Live smartwatch during our experiment. Their upper arm and forearm lengths, l_u and l_f, were measured beforehand.

Experiments with each volunteer were executed in three sessions. In the first session, volunteers were asked to move their arms totally freely for 3 minutes. Users performed random, meaningless, arm gestures – we requested them to not move their hands behind their backs to avoid losing ground truth from the Kinect. In the second session, we deliberately asked the volunteers not to put their elbows above the shoulder. The goal is to mimic real-world scenarios where most of the gestures need the elbow to be at lower heights.

Under this constraint, the volunteers again moved their arms freely for 3 minutes. In the third session, volunteers were asked to repeat a set of pre-defined gestures for 10 times. The gesture set contains eating, drinking, boxing, bouncing a basketball, weight lifting, drawing a circle, drawing a triangle, drawing a square, and writing numeric digits in the air. During the whole experiment, a Kinect 2.0 was placed in front the volunteer to record ground truth. We prevent any movement in the background since that affects Kinect's ground truth calculations.

6.3 Performance Results

The following questions are of our interest in this section:

1. How well can *ArmTrak* track arm postures in general?
2. How does *ArmTrak*'s performance vary among different users and with pre-defined gestures? Are certain gestures better than others?
3. Will error accumulate and *ArmTrak*'s tracking diverge over time?
4. What is the accuracy and latency tradeoff with the real-time version and the offline cloud version?
5. How well can *ArmTrak* track 2D shapes of different objects and digits drawn by users?

In all these results, we measure error for every time step (i.e., output of HMM) and draw the CDF over all measurements.

(1) How well can *ArmTrak* track arm postures?

Figure 6.1 (a) and (b) show the CDF of tracking errors for the elbow and wrist, respectively. The results are for the higher-accuracy HMM version. For free-form motion (where users performed completely random gestures), the median errors for the elbow and wrist are around 7.9 cm and 9.2 cm. Once the elbow was restricted to remain below the shoulder, the error reduced further to 6.6 cm and 8.3 cm for the elbow and wrist, respectively.

Finally, for pre-defined gestures like "eating", "weight lifting", etc., we use the ground truth information from the seven other users as priors for the eighth user, and perform cross-validation. Observe that the median error drops even further to 4.5 cm and 5.7 cm for elbow and wrist on average. Compared to the volunteers' average arm length of 50.2 cm, we believe *ArmTrak*'s accuracy, with minimal prior information, makes it amenable to most gesture recognition applications, including gaming control, TV control, and daily activity patterns such as eating or exercising. Of course, the results can improve appreciably with more application-specific prior information.

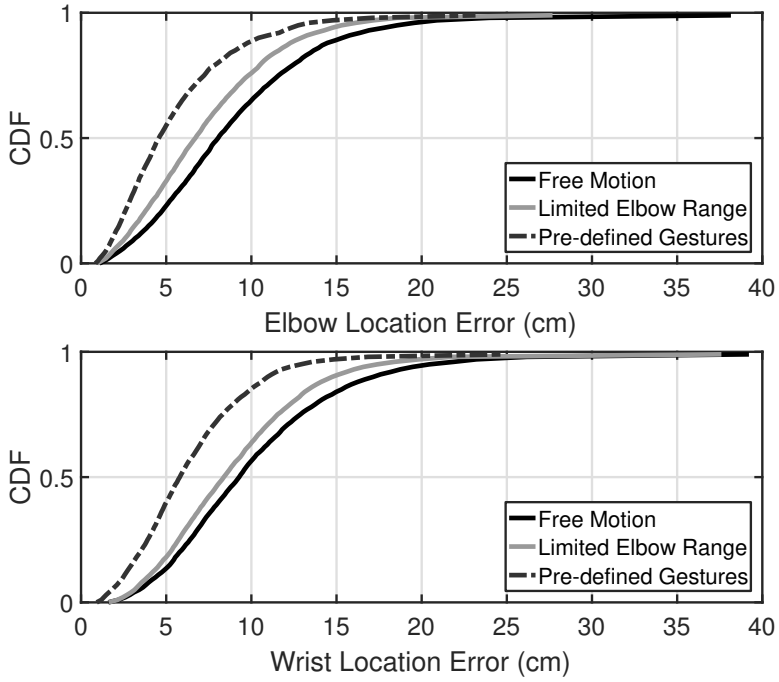

Figure 6.1: (a) *ArmTrak*'s performance on elbow tracking. (b) *ArmTrak*'s performance on wrist tracking.

(2) How does *ArmTrak*'s performance vary across different users and pre-defined gestures?

Figure 6.2 plots the performance across different users. We observe that performance is consistently high across all users, across all the three categories – pre-defined gestures, limited elbow range, and free motion. Also, the trend that pre-defined gestures are generally better than limited elbow range cases,

and limited elbow range better than free-motion cases, holds across the users. The performance is worst for user 7. On examining the Kinect video, we find that user 7 moved that hand extremely fast. Still, the errors were around 10 cm.

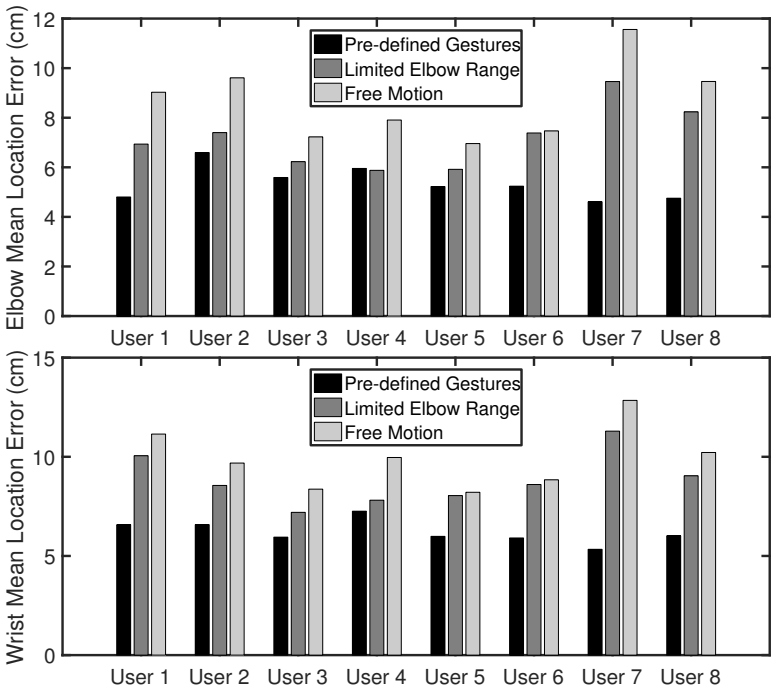

Figure 6.2: *ArmTrak*'s performance on (a) elbow and (b) wrist tracking, for different users.

Figure 6.3 plots *ArmTrak*'s performance across all types of eight gestures. Evidently, the performance did not show major variations across gestures, suggesting that the system is not biased to any patterns. Considering the fact that the prior is only obtained from seven other users, we gained confidence that by improving the prior, *ArmTrak* can be generalized to and work well on other pre-defined gestures.

Upon comparing the performance among these gestures, we find that "eating" has the best performance and "drawing a triangle" is the worst. We again look into the Kinect video data and find that when volunteers were performing "eating" gestures, their elbow only moved in a smaller region, while with "drawing a triangle", the shape, size and position where they drew the triangle are all different, incurring a far greater elbow range.

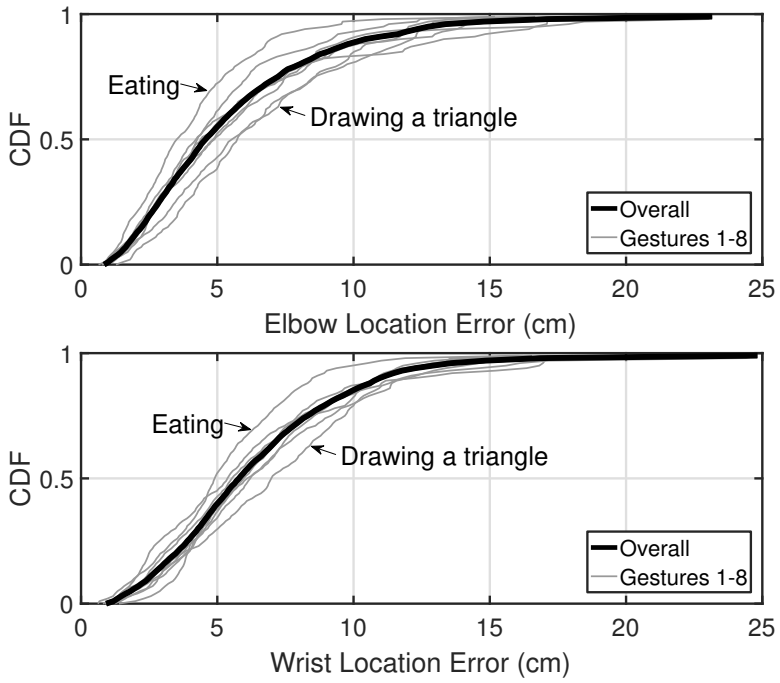

Figure 6.3: *ArmTrak*'s performance on (a) elbow and (b) wrist tracking, for different activities.

(3) Will error accumulate and *ArmTrak*'s performance degrade over time?

ArmTrak attempts to find a sequence of smartwatch locations (from the changing point clouds) that best matches with the observed acceleration data. This global optimization within point clouds ensures that the error will not accumulate over time, as it does with unconstrained double integration. One may argue that this optimization is performed offline over the whole motion sequence, thus intermediate states also benefit from future data. Therefore, characterizing the errors at every intermediate state, with no look-ahead into the future, is also of interest.

To understand the impact, we performed another experiment in which we ask the HMM to traceback to each timestamp and report the instantaneous location estimate at that point. Figure 6.4 shows the general wrist tracking error trend for three volunteers. Although this error is higher than applying the global Viterbi decoding over the whole sequence, the error still does not accumulate. We believe this is perhaps the most important property of *ArmTrak*.

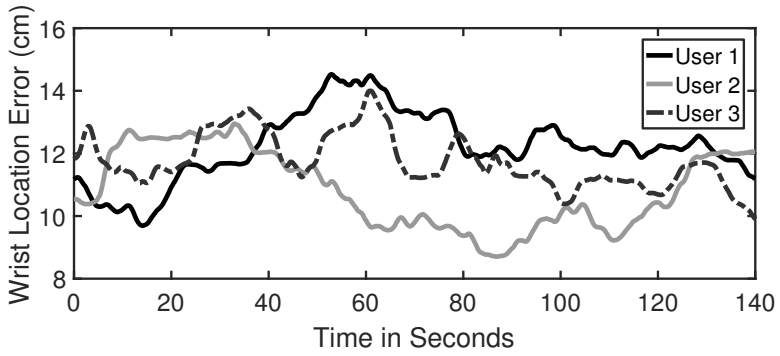

Figure 6.4: The general (smoothed) trend for wrist location error over time.

(4) What is the accuracy and latency tradeoff with the real-time version and the offline cloud version?

Figure 6.5 shows the comparison of tracking accuracy for both elbow and wrist, using real-time and offline algorithms. Recall that the real-time algorithm computes a weighted average of the point cloud and stores a lookup table in the watch, indexed by 3D watch orientation. Compared with the offline algorithm, the median error of the real-time version increases to 12.0 cm for the elbow and 13.3 cm for the wrist. Although high, the real-time version also remains stable over time.

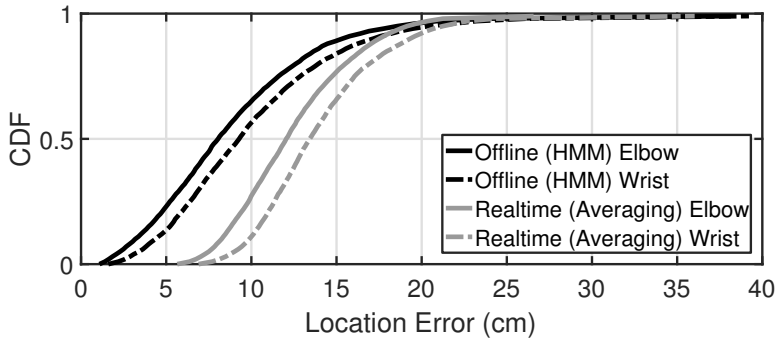

Figure 6.5: Performance comparison between offline (HMM) and real-time (simple averaging).

We computed the delay of both offline and real-time versions of *ArmTrak*. The real-time version running on the watch is essentially a lookup with orientation as index – even for very high update rate the lookup time is negligible. The delay of offline-version contains both the network upload/download delay of the sensor data, plus the computation time at the cloud (i.e., running MATLAB on a quad core graduate student desktop). The results are re-

ported in Table 6.1 for 5 Hz update frequency (i.e., HMM updates five times per second). Evidently, longer gesture data incurs almost a 10x increase in computation.

Table 6.1: Latency of offline version increases with increase in the trace length. This is because Viterbi decoding computes the globally optimal sequence of states, and incurs $O(N^3T)$ complexity.

Trajectory Time	10 s	30 s	1 min	3 min
Delay	98.2 s	289.3 s	9.1 min	26.9 min

(5) What shapes have been inferred by *ArmTrak*?

Figure 6.6 shows some sample trajectories of the wrist, when users were asked to draw shapes and digits in the air. Although the reproduction is not perfect, *ArmTrak* tracks the trend of the trajectory quite well. All users expressed satisfaction when they were shown the shapes that they drew in the air.

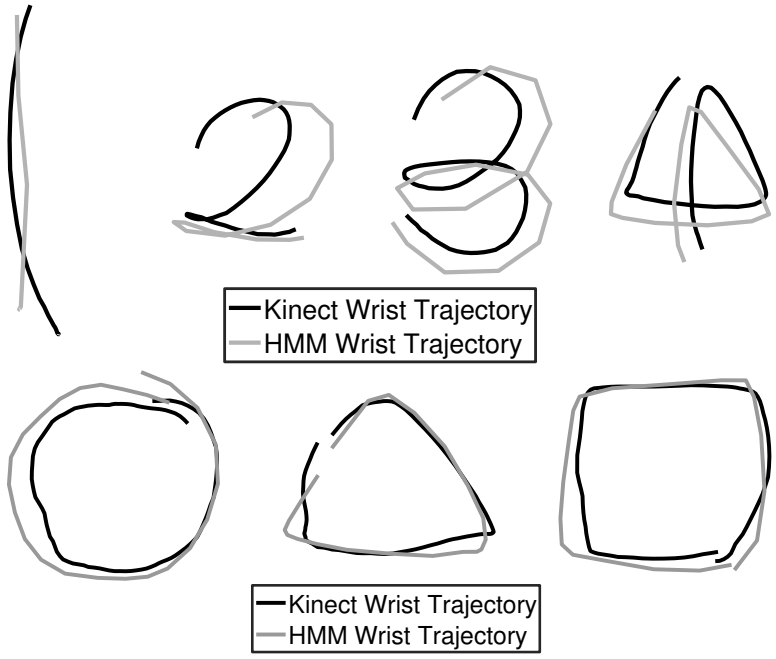

Figure 6.6: *ArmTrak*'s tracking result for (a) writing four digits, and (b) drawing simple shapes.

For more complex situations, we asked the user to draw complicated shapes

like a "star" or an "Olympic ring". Figure 6.7 shows one such case where the user drew for almost 1 minute – *ArmTrak* was consistently able to track the 3D shape.

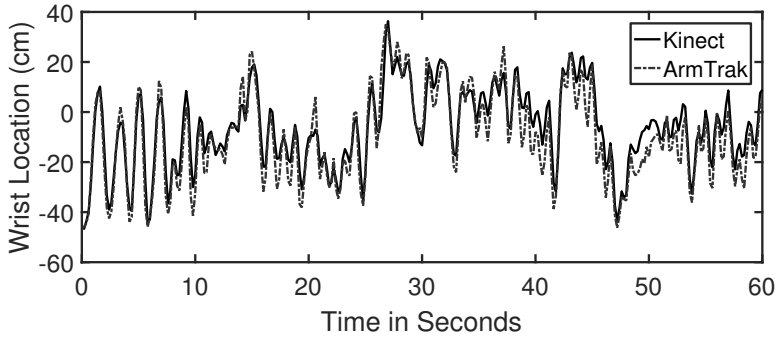

Figure 6.7: *ArmTrak*'s tracking result for a complicated 3D trajectory (only Z-axis is shown).

CHAPTER 7

LIMITATIONS AND NEXT STEP

A few more technical pieces will need to come together before *ArmTrak* can be viewed as a usable technology – we discuss these below.

(1) **Facing Direction and Tracking on the Move**: We opportunistically estimate the user's facing direction when his or her hand is in the vertical free-fall posture, or swinging through this position perhaps while walking. Admittedly, we have not stress-tested this and are not confident this would scale in completely uncontrolled situations. For instance, if a user is sitting for a long duration, the free-fall opportunity may not arise, but the user may change her facing direction (perhaps by swiveling her chair). A deeper treatment of facing direction is necessary to better ground the initial state of the watch. On related lines, *ArmTrak* will falter when the user is on the move – we have side-stepped this case, but plan to evaluate the extent of degradation in future.

(2) **Need for More Speed**: The Viterbi decoding is running on a quad-core student desktop and is roughly producing results at 10x rate. This means that tracking the arm motion for τ seconds requires 10τ seconds of processing time. Of course, more hardware on a cloud can certainly bring down this latency, but the more important question pertains to whether more speedup is possible. We believe some degree of optimization (such as beam search) and parallelism would be possible inside the dynamic programming; we also believe a marginal sacrifice in accuracy can offer considerable speedup. The latency-accuracy tradeoff proved far richer than we anticipated and we intend to investigate this thoroughly in future.

(3) **Energy Consumption**: We have ignored the energy implications of our technique. For the real-time system, we expect the weighted averaging technique to impose minimal energy burden. For the HMM/Viterbi model, we expect the system to run on the cloud – thus the energy consumption mainly emerges from network uploads. We have not characterized this over-

head, however, it appears that many offloading applications are viable under this model. This is perhaps because most applications are likely to be on-demand (e.g., the user turns on the app during the visit to the gym and turns off thereafter). Further, some apps can tolerate latency and can delay the uploads until the watch is connected to power.

CHAPTER 8

CONCLUSION

This thesis demonstrates our attempt to estimate/track the geometric motion of the human arm, using only the inertial sensors on the smartwatch. The problem is challenging primarily because the smartwatch is a single point of measurement on this otherwise large space of possibilities. Moreover, the measurements are noisy, making continuous tracking over longer time scales even more difficult. We develop *ArmTrak*, a system that distills observations from human kinematics, and uses them carefully inside a (modified) HMM framework. Our results are encouraging, and with more effort, could become a useful underlay to a broad class of gesture-based applications.

REFERENCES

[1] A. Parate, M.-C. Chiu, C. Chadowitz, D. Ganesan, and E. Kalogerakis, "RisQ: Recognizing smoking gestures with inertial sensors on a wristband," in *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2014, pp. 149–161.

[2] Y. Dong, A. Hoover, J. Scisco, and E. Muth, "A new method for measuring meal intake in humans via automated wrist motion tracking," *Applied Psychophysiology and Biofeedback*, vol. 37, no. 3, pp. 205–215, 2012.

[3] H. Wang, T. T.-T. Lai, and R. Roy Choudhury, "Mole: Motion leaks through smartwatch sensors," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 155–166.

[4] C. Xu, P. H. Pathak, and P. Mohapatra, "Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch," in *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*. ACM, 2015, pp. 9–14.

[5] "Rithmio," http://rithmio.com/.

[6] D. Roetenberg, H. Luinge, and P. Slycke, "Xsens MVN: full 6DOF human motion tracking using miniature inertial sensors," *Xsens Motion Technologies BV, Tech. Rep*, 2009.

[7] J. O'Rourke, N. Badler et al., "Model-based image analysis of human motion using constraint propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 6, pp. 522–536, 1980.

[8] D. Hogg, "Model-based vision: A program to see a walking person," *Image and Vision Computing*, vol. 1, no. 1, pp. 5–20, 1983.

[9] K. Rohr, "Towards model-based recognition of human movements in image sequences," *CVGIP: Image Understanding*, vol. 59, no. 1, pp. 94–115, 1994.

[10] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55–79, 2005.

[11] V. K. Singh, R. Nevatia, and C. Huang, "Efficient inference with multiple heterogeneous part detectors for human pose estimation," in *European Conference on Computer Vision*. Springer, 2010, pp. 314–327.

[12] X. Lan and D. P. Huttenlocher, "Beyond trees: Common-factor models for 2d human pose recovery," in *IEEE International Conference on Computer Vision*, vol. 1. IEEE, 2005, pp. 470–477.

[13] D. Ramanan and C. Sminchisescu, "Training deformable models for localization," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2006, pp. 206–213.

[14] B. Sapp, A. Toshev, and B. Taskar, "Cascaded models for articulated pose estimation," *Computer Vision–ECCV 2010*, pp. 406–420, 2010.

[15] M. P. Kumar, A. Zisserman, and P. H. Torr, "Efficient discriminative learning of parts-based models," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 552–559.

[16] M. Andriluka, S. Roth, and B. Schiele, "Pictorial structures revisited: People detection and articulated pose estimation," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 1014–1021.

[17] Y. Yang and D. Ramanan, "Articulated pose estimation with flexible mixtures-of-parts," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1385–1392.

[18] "Microsoft kinect," https://dev.windows.com/en-us/kinect.

[19] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images," *Communications of the ACM*, vol. 56, no. 1, pp. 116–124, 2013.

[20] H. Zhou, H. Hu, and Y. Tao, "Inertial measurements of upper limb motion," *Medical and Biological Engineering and Computing*, vol. 44, no. 6, pp. 479–487, 2006.

[21] H. Zhou and H. Hu, "Upper limb motion estimation from inertial measurements," *International Journal of Information Technology*, vol. 13, no. 1, pp. 1–14, 2007.

[22] A. G. Cutti, A. Giovanardi, L. Rocchi, A. Davalli, and R. Sacchetti, "Ambulatory measurement of shoulder and elbow kinematics through inertial and magnetic sensors," *Medical & Biological Engineering & Computing*, vol. 2, no. 46, pp. 169–178, 2008.

[23] M. El-Gohary and J. McNames, "Shoulder and elbow joint angle tracking with inertial sensors," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 9, pp. 2635–2641, 2012.

[24] S. Shen, H. Wang, and R. Roy Choudhury, "I am a smartwatch and I can track my user's arm," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, 2016.

[25] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, no. 2, pp. 90–126, 2006.

[26] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 4, pp. 743–761, 2012.

[27] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern Recognition*, vol. 36, no. 3, pp. 585–601, 2003.

[28] N. Robertson and I. Reid, "A general method for human activity recognition in video," *Computer Vision and Image Understanding*, vol. 104, no. 2, pp. 232–248, 2006.

[29] P. C. Ribeiro and J. Santos-Victor, "Human activity recognition from video: Modeling, feature selection and classification architecture," in *Proceedings of International Workshop on Human Activity Recognition and Modelling*. Citeseer, 2005, pp. 61–78.

[30] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *Communications Surveys & Tutorials, IEEE*, vol. 15, no. 3, pp. 1192–1209, 2013.

[31] "Fitbit," https://www.fitbit.com/.

[32] "Apple Watch," http://www.apple.com/watch/.

[33] J. Tautges, A. Zinke, B. Krüger, J. Baumann, A. Weber, T. Helten, M. Müller, H.-P. Seidel, and B. Eberhardt, "Motion reconstruction using sparse accelerometer data," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 3, p. 18, 2011.

[34] Q. Riaz, G. Tao, B. Krüger, and A. Weber, "Motion reconstruction using very few accelerometers and ground contacts," *Graphical Models*, vol. 79, pp. 23–38, 2015.

[35] H. Zhou and H. Hu, "Inertial motion tracking of human arm movements in stroke rehabilitation," in *IEEE International Conference Mechatronics and Automation, 2005*, vol. 3. IEEE, 2005, pp. 1306–1311.

[36] "Motion capture systems - vicon," http://vicon.com/.

[37] "Motion capture systems - optitrack," http://optitrack.com/.

[38] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3d tracking via body radio reflections," in *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, 2014, pp. 317–329.

[39] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*. ACM, 2013, pp. 27–38.

[40] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand, "Capturing the human figure through a wall," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, p. 219, 2015.

[41] T. Li, Q. Liu, and X. Zhou, "Practical human sensing in the light," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2016, pp. 71–84.

[42] R. L. Gajdosik and R. W. Bohannon, "Clinical measurement of range of motion review of goniometry emphasizing reliability and validity," *Physical Therapy*, vol. 67, no. 12, pp. 1867–1872, 1987.

[43] D. Boone and S. Azen, "Normal range of motion of joints in male subjects." *The Journal of bone and joint surgery. American volume*, vol. 61, no. 5, pp. 756–759, 1979.

[44] J. C. Perry and J. Rosen, "Design of a 7 degree-of-freedom upper-limb powered exoskeleton," in *The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics*. IEEE, 2006, pp. 805–810.

[45] R. S. Hartenberg and J. Denavit, "A kinematic notation for lower pair mechanisms based on matrices," *Journal of Applied Mechanics*, vol. 77, no. 2, pp. 215–221, 1955.

[46] N. B. Reese and W. D. Bandy, *Joint Range of Motion and Muscle Length Testing*. Elsevier Health Sciences, 2013.

[47] P. Zhou, M. Li, and G. Shen, "Use it free: Instantly knowing your phone attitude," in *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*. ACM, 2014, pp. 605–616.